Specification

Crystal of glucokinase protein and drug design method using crystal thereof.

The Field of Technology

This invention relates to crystal of novel glucokinase protein (hereinafter it is called "GK protein") and the drug design method or the like employing three-dimensional structure coordinates obtained by using the crystal thereof.

Background Technique

Glucokinase (ATP: D-hexose 6-phosphotransferaze, EC2.7.1.1) is one of four kinds of hexokinase isozymes of mammals (hexokinase IV). These isozymes catalyse the same reaction, however, differences exist in the Km value with respect to glucose. In other words, the Km value of hexokinase I, II and III being 10^{-6}-10^{-4} M, but on the other hand the Km value of hexokinase IV, called glucokinase with respect to glucose is much greater at about 10^{-2} M. Hexokinase is an enzyme participating in the initial stage of glycolytic pathway, and catalyses the reaction from the glucose to glucose-6-phosphate.

As for glucokinase, the expression is mainly localised in liver and pancreatic beta cell, and it plays an important role in glucose metabolism of the whole body by controlling the rate-determining step of glucose metabolism in these cells. As for the glucokinase of liver and pancreatic beta cell, the sequence of 15 amino acids at N terminal is respectively different due to splicing difference. However, the enzymatic characteristics are the same.

The hypothesis that glucokinase acts as glucose sensor of pancreatic beta cell and liver is proposed since approximately 10 years ago (Garfinkel D, et al: Am J Physiol 247 [3Pt2]: R527-536, 1984). In practice, it is becoming clear from results of recent glucokinase gene manipulation mouse, that the glucokinase plays an important role in glucose homeostasis of the whole body.

The mouse in which glucokinase gene is destroyed dies of diabetes mellitus soon after birth (Grupe A, et al: Cell 83: 69-78. 1995). On the other hand, as for the mouse which overexpressed glucokinase, the blood glucose level becomes low (Ferre T, et al: Proc Natl Acad Sci USA 93: 7225-7230. 1996). When glucokinase activity is increased by the rise in glucose concentration, although the reactions of

pancreatic beta cell and hepatocyte are different, in each case, it acts in the direction of lowering blood glucose. The pancreatic beta cell starts to secrete more insulin, the liver takes up sugar and stores as glycogen and at the same time lowers the sugar release.

In this way, the fluctuation of glucokinase enzyme activity plays an important role in glucose homeostasis of mammal through liver and pancreatic beta cell. Glucokinase gene mutation is discovered in the case that develops diabetes mellitus in youth known as MODY2 (maturity-onset diabetes of the young), and the lowering of glucokinase activity is said to be the cause of blood glucose rise (Vionnet N, et al.: Nature 356: 721-722, 1992). On the other hand, the lineage having mutation to increase glucokinase activity is also found, and such persons show hypoglycemic symptom (Glaser B, et al.: N Engl J Med 338: 226-230, 1998).

From the above, glucokinase also acts glucose sensor in human and plays an important role in glucose homeostasis. On the other hand, because the glucokinase of many type II diabetics is not mutated, the blood glucose control using glucokinase sensor system is considered possible. Because the glucokinase activator substance can be expected to have insulin secretion promotion action of pancreatic beta cell and sugar up take acceleration and sugar release suppression actions in liver, it is considered as useful therapeutic drug of type II diabetic patients.

Recently, a localised expression of pancreatic beta cell type glucokinase was found in rat brain, in particular in ventromedial hypothalamic nucleus (Ventromedial hypothalamus, VMH) which is the feeding centre. About 20 % of neurons of VMH was known as glucose responsive neuron, and it has been considered to play an important role in weight control in the past. When glucose is administered intracerebrally to rat, food consumption falls, whereas when the glucose metabolism is suppressed by administration of glucose analogue, glucosamine in brain, overeating occurs. From electrophysiological experiment, glucose responsive neurons are found to be activated in response to physiological glucose concentration changes (5-20 mM), however its activity is suppressed when the glucose metabolism is inhibited with glucosamine and the like. In glucose concentration sensing system of VMH, a mechanism through glucokinase the same as insulin secretion of pancreatic beta cells is assumed. Accordingly, there is a possibility that a substance that causes glucokinase activation of VHM in addition to liver, pancreatic beta cell is expected to correct problem of obesity which is a problem in many type II diabetic patients,

in addition to blood glucose correction effect.

On the other hand, it is described in DIABETES, vol. 48, 1698-1705, September 1999 that the stereostructure of glucokinase was predicted from hexokinase 1. However, in practice, crystallisation was not carried out, nor it was a practical one.

In accordance with the above, to elucidate three-dimensional stereostructure of glucokinase and to enable efficient discovery of a compound that interacts with glucokinase are thought to greatly contribute to the development of for example a therapeutic agent or preventative agent of diabetes, a therapeutic agent or preventative agent of chronic complication of diabetes mellitus such as retinopathy, nephropathy, neurosis, ischemic cardiac disease, arteriosclerosis or the like, a therapeutic agent or preventative agent of obesity.

Presently, CARDD (Computer Aided Rational Drug Design) using computer for the tasks such as analysis of active centre of a protein or a prediction of reaction mechanism has been employed at practical level.

In the drug creation system using CARDD, the structure of active site of protein is predicted based on the three-dimensional structure analysis data of the target protein. And information about candidate compounds which can bind to the structure of active site thereof is obtained from the compound database. Thereafter, on consideration of the three-dimensional structure and physical properties of the active site of the target protein and the candidate compound, candidate compounds which can bind to the target protein are selected. These steps are so-called in silico screening step.

Whether the compound selected by in silico screening step binds to the target protein and change the activity thereof or not, is examined by actual examination (wet experiment). And the compound that changes the activity of the target protein becomes the effective ingredient of a drug. Thereby a compound that interacts with the target protein can be efficiently screened without carrying out the procedure wherein innumerable compounds are acted on the target protein one by one and the interactions are confirmed.

In silico screening can be said as an effective means of pharmaceutical development because the

candidate compounds that bind to the target protein can be greatly narrowed down.

Three-dimensional structure analysis data by X-ray structure analysis of the target protein becomes an important information in drug creation system using CARDD. Crystal of target protein is required as analysis sample in three-dimensional structural analysis by X-ray structure analysis. Accordingly, in order to carry out development of drug creation related to GK based on the drug creation system using CARDD, the crystal of GK is required. However, as stated above, crystallisation of GK was difficult, and it could not provide information necessary for CARDD.

This invention was made on consideration of the problems of aforesaid technology of the prior art, and had objects to obtain crystal of glucokinase and to design compounds that bind to glucokinase based on the information obtained from aforesaid crystal.

Disclosure of the Invention

At least one of aforesaid objects is solved by the following invention.

[1] A glucokinase protein characterised in being used for crystallisation.

[2] A protein in accordance with aforesaid [1] comprising amino acid sequence in accordance with Sequence Number 5.

[3] A crystal of protein comprising amino acid sequence in accordance with Sequence Number 5 or amino acid sequence substantially the same amino acid sequence thereof.

[4] A crystal in accordance with aforesaid [3], wherein the said protein is glucokinase protein.

[5] A crystal in accordance with aforesaid [3] comprising crystals of protein containing amino acid sequence in accordance with Sequence Number 5.

[6] A crystal in accordance with aforesaid [3], wherein the lattice constant satisfies the following equations (1)-(4)

a = b = 79.9 +/- 4 Å (1)

c = 322.2 +/- 15 Å (2)

alpha = beta = 90° (3)

gamma = 120° (4)

[7] A crystal in accordance with aforesaid [6], wherein the space group is $P6_522$.

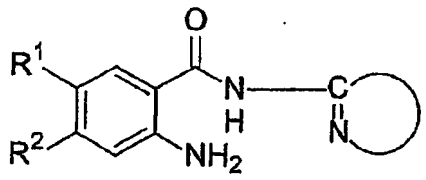
[8] A crystal of protein specified by three-dimensional structure coordinates data in accordance with Table 1.

[9] A crystal wherein in three-dimensional structure coordinates data changed in at least one data of three-dimensional structure coordinates data in accordance with Table 1, the mean square error between atoms of main chain of amino acid represented by three-dimensional structure coordinates data in accordance with Table 1 (C alpha atom) and C alpha atoms represented by the said changed three-dimensional structure coordinates data corresponding to aforesaid C alpha atoms is 0.6 Å or less.

[10] A crystal in accordance with any of [3]-[9], wherein the compound binding site is constructed by at least one of amino acid residues of tyrosine 61 - serine 69, glutamic acid 96 - glutamine 98, isoleucine 159, methionine 210 - tyrosine 215, histidine 218 - glutamic acid 221, methionine 235, arginine 250, leucine 451 - lysine 459 in amino acid sequence shown in sequence Number 5.

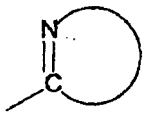
[11] A crystal including a complex of the protein comprising amino acid sequence in accordance with Sequence Number 5 or amino acid sequence substantially the same amino acid sequence thereof and a compound which can bind to the said protein.

[12] A crystal in accordance with aforesaid [11], wherein aforesaid compound is represented by formula (1).



(I)

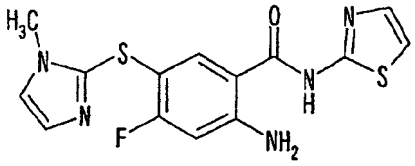
[wherein, R1 shows halogen atom, -S-(O)p-A, -S-(O)q-B or -O-B (wherein, p and q are the same or different and denote an integer of 0-2, A denotes C1-C6 alkyl group of optionally substituted straight chain, B denotes optionally substituted five-membered or six-membered ring aryl group or heteroaryl group, R2 denotes a hydrogen atom or halogen atom, and



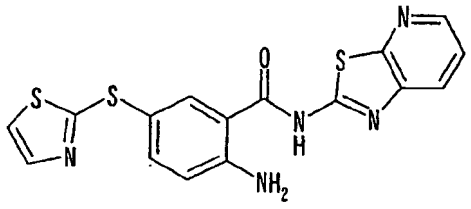
(II)

denotes an optionally substituted monocyclic or bicyclic heteroaryl group having a nitrogen atom adjacent to the carbon atom bonded to amide group].

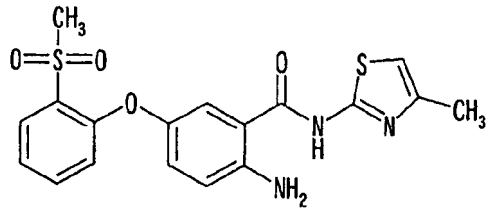
[13]. A crystal in accordance with aforesaid [12], wherein aforesaid compound is any of the compound represented by formula (IIIa)-(IIIc).



(IIIa)



(IIIb)



(IIIc)

[14] A protein in accordance with aforesaid [1] comprising amino acid sequence in accordance with Sequence Number 8.

[15] A crystal of protein comprising amino acid sequence in accordance with Sequence Number 8 or amino acid sequence substantially the same amino acid sequence thereof.

[16] A crystal in accordance with aforesaid [15], wherein the said protein is glucokinase protein.

[17] A crystal in accordance with aforesaid [15] comprising crystals of protein containing amino acid sequence in accordance with Sequence Number 8.

[18] A crystal in accordance with aforesaid [15], wherein the lattice constant satisfies the following equations

a = b = 103.2 +/- 5 Å (5)

c = 281.0 +/- 7 Å (6)

alpha = beta = 90° (7)

gamma = 120° (8)

[19] A crystal in accordance with aforesaid [18], wherein the space group is $P6_522$.

[20] A crystal of protein specified by three-dimensional structure coordinates data in accordance with Table 2.

[21] A crystal wherein in three-dimensional structure coordinates data changed at least one data of three-dimensional structure coordinates data in accordance with Table 2, the mean square error between atoms

of main chain of amino acid represented by three-dimensional structure coordinates data in accordance with Table 2 (C alpha atom) and C alpha atoms represented by the said changed three-dimensional structure coordinates data corresponding to aforesaid C alpha atoms is 0.6 Å or less.

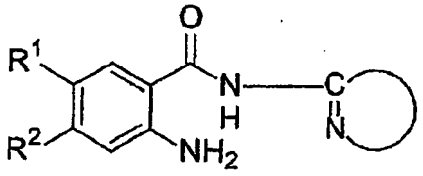
[22] A process for the production of crystal containing a complex of protein and a compound that binds to the protein thereof, including

a protein production step wherein a protein containing the amino acid sequence having deletion of 1-50 amino acid residues from either or both of N terminal and C terminal of the protein containing amino acid sequence in accordance with Sequence Number 2 is produced, and

a protein reaction step wherein a compound that binds to the protein obtained in the said protein production step and the protein obtained in the said protein production step are reacted.

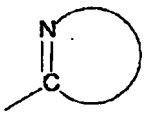
[23] A process to produce crystal of the kind wherein a crystal of a protein is produced, characterised in that a protein including amino acid sequence in accordance with Sequence Number 5 or amino acid sequence substantially the same amino acid sequence thereof and having glucokinase activity and a compound which can bind to the said protein are used.

[24] A process for the production of crystalline protein in accordance with aforesaid [23], wherein the compound which can bind to said protein is a compound represented by formula (1).



(I)

[wherein, R1 shows halogen atom, -S-(O)p-A, -S-(O)q-B or -O-B (wherein, p and q are the same or different and denote an integer of 0-2, A denotes C1-C6 alkyl group of optionally substituted straight chain, B denotes optionally substituted five-membered or six-membered ring aryl group or heteroaryl group, R2 denotes a hydrogen atom or halogen atom, and



(II)

denotes an optionally substituted monocyclic or bicyclic heteroaryl group containing nitrogen atom adjacent to the carbon atom bonded to amide group].

[25]. A process for the production of crystal in accordance with aforesaid [23] or [24] using co-crystallisation or soaking method

[26] A drug design method of the kind wherein based on stereostructural information of a protein, the structure of compound that binds to said protein is designed, characterised in that the stereostructure information of said protein is the information obtained by analysing crystal in accordance with any one of aforesaid [3]-[13] or [15]-[21].

[27] A drug design method in accordance with aforesaid [26] characterised in that
a binding site deduction step wherein the compound binding site of said protein is deduced based on aforesaid stereostructure information, and
a selection step wherein a compound compatible to the compound binding site deduced in aforesaid binding site deduction step is selected from compound library,
are included.

[28] A drug design method in accordance with aforesaid [26] characterised in that
a binding site deduction step wherein the compound binding site of said protein is deduced based on aforesaid stereostructure information, and
a compound structure assembly step wherein the structure of compound compatible to compound binding site deduced in aforesaid binding site deduction step is constructed,
are included.

[29] A drug design method in accordance with aforesaid [26] characterised in that

a binding site deduction step wherein the compound binding site of said protein is deduced based on aforesaid stereostructure information, and

a design step wherein the structure of compound is designed by visual observation so that the compound binding site deduced in aforesaid binding site deduction step and a compound compatible to said compound binding site interact,

are included.

[30] A drug design method in accordance with any of aforesaid [26]-[29], wherein aforesaid compound binding site is constituted by at least one of amino acid residue of tyrosine 61 - serine 69, glutamic acid 96 - glutamine 98, isoleucine 159, methionine 210 - tyrosine 215, histidine 218 - glutamic acid 221, methionine 235, arginine 250, leucine 451 - lysine 459 in amino acid sequence shown in sequence Number 5.

[31] A drug design method in accordance with any of aforesaid [26]-[30] further including a step to measure physiological activity of the candidate compound predicted to be compatible to aforesaid compound binding site.

[32] A drug design method in accordance with any of aforesaid [26]-[30] further including a binding determination step wherein the candidate compound predicted to be compatible to aforesaid compound binding site and a protein including amino acid sequence in accordance with and Sequence Number 5 or amino acid sequence which is substantially the same amino acid sequence thereof are contacted, and whether the candidate compound binds to the said protein or not is assessed.

[33] A process for the production of compound array including the compound group selected by drug design method in accordance with any of aforesaid [26]-[30] is combined as compound array.

Brief Description of the Figures

Figure 1 is a ribbon diagram showing three-dimensional structure of glucokinase.

(Figure 1a is a ribbon diagram showing the structure of glucokinase (Δ1-11)/glucose/compound 1 (compound of formula IIIa). Moreover, the figure on the right is a rotated figure of the figure on the left.)

(Figure 1b is a ribbon diagram showing the simple substance of glucokinase [ΔI-15]. Moreover, the figure on the right is a rotated figure of the figure on the left.)

Figure 2 is a figure showing coupling scheme of compound 1 (compound of formula IIIa) with respect to the binding site of glucokinase (ΔI-11)

Figure 3 is a figure showing the structure of binding site of glucokinase (ΔI-11).

Ideal form for Carrying Out the Invention

In this specification, amino acids, peptides and proteins are represented using abbreviations adopted from the IUPAC-IUB biochemistry designation committee (CBN) shown below. Moreover, the sequence of amino acid residues of peptide and protein are represented so that the N terminal to C terminal comprises from the left end to the right end and moreover the N terminal comprises the first.

Hereinafter, each embodiment of this invention is described in greater detail.

(Glucokinase protein).

Firstly, this invention puts forward glucokinase protein characterised in being used for crystallisation. Glucokinase protein (GK protein) is involved in extremely important sugar metabolism in vivo as described above. Accordingly, by solving the three-dimensional structure of GK protein and by elucidating active site of GK protein, it is possible to search compounds that bind to GK protein (activator or inhibitor). Therefore it is important to clarify the three-dimensional structure of GK protein.

As technique to clarify the three-dimensional structure of protein, X-ray crystal structure analysis is well known. In other words, protein is crystallised, mono-chromatised X-ray is irradiated to the crystal thereof, and three-dimensional structure of said protein is elucidated on the basis of the obtained X-ray diffraction pattern (Blundell, T.L. and Johnson, L.N, PROTEIN CRYSTALLOGRAPHY, pp. 1-565, (1976) Academic Press, New York). First GK protein needs to be crystallised in order to provide for the x-ray crystal structure analysis of GK protein.

Wherein, the "GK protein" of this invention refers to human derived liver type glucokinase having amino acid sequence shown in sequence Number 2 and a protein containing amino acid sequence which is substantially the same as Sequence Number 2. Wherein, as aforesaid protein containing amino acid sequence which is substantially the same, a species having glucokinase activity is preferable. Accordingly, in this specification, the GK protein includes not only the human derived liver type glucokinase, however also human derived pancreas type glucokinase, and non-human derived GK proteins such as mouse, rat, monkey and the like. In this invention human liver type glucokinase is preferably used. In glucokinase derived from human, 15 amino acid residues at N terminal differ in the liver type and the pancreas type. Wherein, "glucokinase activity" refers to an activity to catalyse reaction from glucose to glucose-6-phosphate.

It is generally well known that the crystallisation of protein is difficult, and the GK protein was not able to be crystallised without treatment. These inventors carried out various investigations with trial and error, as a result succeeded in crystallisation of GK protein for the first time by deletion of 11 or 15 amino acids at the N terminal side of GK protein. It was thought that the deleted region protruded from the globular GK protein molecule when the crystallisation was attempted, as a result, caused steric hindrance between adjacent GK protein molecules, and prevented the crystallisation of the GK protein. In other words, in this invention, the crystal of GK protein was obtained by using a GK protein in which 11 amino acid residues at N terminal side is deleted (Sequence Number 5) or a GK protein in which 15 amino acid residues at N terminal side is deleted (Sequence Number 8) in the glucokinase in which amino acids sequence had been known however the crystallisation had been unsuccessful. Wherein the number of amino acids is not restricted as long as it is within a range that the steric hindrance disappears between adjacent crystals. In an embodiment for example, in amino acid sequence represented by Sequence Number 2, amino acids sequence or the like in which amino acid residues of 1-50, preferably 3-30, more preferably 5-25, more preferably still 8-18, most preferably 11-15 at N terminal side are deleted, can be used in this invention. Moreover, the amino acid sequence or the like in which amino acid residues of 1-8, preferably 1-7, more preferably 2-6 at C terminal side are deleted, is used in this invention.

(Crystal of glucokinase protein and a process for the production thereof).

Next, in this invention, crystals including protein containing amino acids sequence in accordance with Sequence Number 5, and Sequence Number 8 or amino acids sequence which is substantially the same as

amino acid sequence thereof are put forward.

As described earlier, as GK protein used in crystallisation, proteins containing amino acids sequence in accordance with Sequence Number 5, and/or Sequence Number 8 or amino acids sequence which is substantially the same as amino acid sequence thereof, or the like are used.

The proteins containing amino acids sequence in accordance with Sequence Number 5, and/or Sequence Number 8 or amino acids sequence which is substantially the same as the amino acid sequence thereof (hereinafter it may be abbreviated as "GK protein" together with proteins containing amino acids sequence in accordance with Sequence Number 2 or amino acids sequence which is substantially the same as the amino acid sequence thereof) can be any as long as crystallisation is possible, and the amino acid sequence thereof is not restricted in particular. Wherein, the proteins containing amino acids sequence which is substantially the same as the amino acid sequence in accordance with Sequence Number 5, and/or Sequence Number 8 does not necessarily have glucokinase activity, and may be an inactive mutant (for example, a mutant inactivated by the presence of mutation at ATP binding site) as long as it has a crystal structure from which the information necessary for drug design can be obtained. Wherein, as proteins containing amino acids sequence which is substantially the same as the amino acid sequence in accordance with Sequence Number 2 or Sequence Number 5, amino acids sequences having about 60 % or more, preferably about 70% or more, more preferably about 80% or more, in particular preferably about 90% or more, and most preferably about 95% or more homology to the amino acids sequence in accordance with Sequence Number 2 or Sequence Number 5, or the like are nominated. Moreover, as proteins containing amino acids sequence which is substantially the same as the amino acids sequence in accordance with Sequence Number 2 or Sequence Number 5, for example, amino acids sequences in which amino acid residues of 1-10, preferably 1-5, more preferably 1-3 more preferably still 1-2 are substituted, deleted, added and/or inserted in the amino acids sequence in accordance with Sequence Number 2 or Sequence Number 5 are exemplified.

Three-dimensional structural analysis of GK protein is carried out for example as follows. Firstly, the protein is purified. And a series of steps such as crystallisation, X-ray diffraction intensity data collection, phase determination of each diffraction spot, electron density calculation, molecular model construction, refinement of structure or the like is carried out. As main equipment for performing

protein structure analysis, incubator for crystallisation, binocular microscope, X-ray diffractometer, three dimensional computer graphics apparatus or the like are used. The actual experimental process to produce protein crystals is divided into step to purify protein in large amount (several mg or more is preferred), a step to widely search conditions for obtaining crystal and a step to obtain high quality crystal suitable for X-ray analysis. Hereinafter, each step is described in concrete terms.

For crystallisation, GK protein is purified to high purity. As process for purification, well known processes can be used, and for example, column chromatography, salt precipitation, centrifugation or the like are used.

Purified GK protein is crystallised and provided as a sample for X-ray crystal structure analysis. Crystallisation is performed based on well known method such as vapor diffusion method, dialysis or the like. When obtaining protein crystals, many elements such as purity / concentration of protein, temperature, pH, concentration of the precipitant used need to be examined. Investigation of crystallisation conditions can be carried out over a wide range using commercial screening reagent, and it is preferably screened using 1-2 μl of protein solution in protein concentration of 1-2 % per condition. In this way when microcrystals or the like are obtained, it is preferred to further refined the conditions.

Moreover, extremely many conditions must be searched in order to obtain crystal of GK protein. Accordingly, a large quantity expression system of the protein is preferably constructed also for the investigation of crystallisation condition. Generally, among proteins, many of the crystallising species are monodispersed in solution state, and polydispersed species do not crystallise in most cases. Therefore, N terminal of GK protein is successively removed, monodispersion properties of protein solution are assessed for the obtained protein using light scattering apparatus, and whether sample is suitable for crystallisation or not may be examined.

Next, using the obtained crystal of GK protein, X-ray diffraction intensity measurement is carried out. Recently, a method wherein the crystal is scooped with a ring of narrow thread or the like, is rapidly cooled to liquid nitrogen temperature, and is measured at low temperature as it is, may also be used. Usually, the intensity measurement of diffracted x-ray is performed by two-dimensional detector such as image plate or the like. Many diffraction lines generated by rotating crystal while irradiating the X-ray

are recorded on image plate, and the recorded diffraction intensities are read by shining a laser.

Next, it is preferred to prepare heavy atom iso-form replacement bodies by heavy atom soaking method or co-crystallisation method. Using this, the phase of the protein crystal can be determined by multiple isomorphous replacement method (MIR method). Instead of introducing heavy atom, the phase is also determined by multiwavelength anomalous scattering method (MAD method) based on the diffraction intensity data using complex X-rays. Molecular replacement method (MR method) in which, when a structure of molecule containing analogous structure has been already solved, the initial structure can be obtained by applying the molecular structure thereof in the crystal, Furrier synthesis diagram is drawn on the basis of this, and the structure of remaining part is elucidated, and the total structure is determined, is known as well.

Once the phase was determined by aforesaid process, electron density is determined from this. The precision of this depends on the number of reflection (resolution) and the precision of the reflection used. The resolution is expressed with the minimum plane spacing of the reflection used. Molecular model is constructed from this electron density diagram. When the molecular model is constructed, the atomic coordinates are obtained, therefore, the calculated value of structure factor is determined from this, and refinement of atomic parameters is carried out by the least-square method to approximate this size to the observed value. In this way, the most reasonable structural information is obtained.

In accordance with this invention, the crystal of GK protein shown in sequence Number 5 has been successfully prepared (cf. later described Example). The obtained crystal of GK protein had lattice constant which satisfied the following equations (1)-(4).

a = b = 79.9 +/- 4 Å (1)

c = 322.2 +/- 15 Å (2)

alpha = beta = 90° (3)

gamma = 120° (4)

Moreover, this crystal was elucidated to have space group $P6_522$. Wherein, aforesaid a = b is preferably 79.9 +/- 3 Å, more preferably 79.9 +/- 2 Å and even more preferably 79.9 +/- 1 Å. Moreover, aforesaid c is preferably 322.2 +/- 10 Å, more preferably 322.2 +/- 8 Å, and even more preferably 322.2 +/- 5 Å.

The three-dimensional structural coordinates of the GK protein crystal obtained in this way are shown in Table 1.

Table 1

Moreover, Table 1 is constructed in accordance with representation method of protein data bank generally used by a person skilled in the art. The GLC in Table 1 denotes glucose molecule, and CP1 denotes the compound represented by formula IIIa, and HOH denotes water molecule.

Moreover, in this invention, the crystal of GK protein shown in sequence Number 8 has been successfully prepared (cf. later described Example). The obtained crystal of GK protein had lattice constant which satisfied the following equations (5)-(8).

a = b = 103.2 +/- 5 Å (5)

c = 281.0 +/- 7 Å (6)

alpha = beta = 90° (7)

gamma = 120° (8)

Moreover, this crystal was elucidated to have space group $P6_522$. Wherein, aforesaid a = b is preferably 103.2 +/- 3 Å, more preferably 103.2 +/- 2 , and even more preferably 103.2 +/- 1 Å. Moreover, aforesaid c is preferably 281.0 +/- 6 Å, more preferably 281.0 +/- 4 Å, and even more preferably 281.0 +/- 2 Å.

The three-dimensional structural coordinates of the GK protein crystal obtained in this way are shown in Table 2.

Table 2

Moreover, Table 2 is constructed in accordance with representation method of protein data bank generally used by a person skilled in the art. The HOH in Table 2 denotes water molecule.

In this invention, crystals of the protein having amino acids sequence which is substantially the same as Sequence Number 5 and/or Sequence Number 8 and having glucokinase activity are within the range of

this invention. As such crystals, for example, crystals wherein in three-dimensional structure coordinates data changed at least one data of three-dimensional structure coordinates data in accordance with Table 1 and/or 2, the mean square error between atoms of main chain of amino acid represented by three-dimensional structure coordinates data in accordance with Table 1 and/or 2 (C alpha atom) and C alpha atoms represented by the said changed three-dimensional structure coordinates data corresponding to aforesaid C alpha atoms is 0.6 Å or less, are nominated. Even if the numerical values of coordinates representing the position of atoms differ, two structural coordinates which can superimpose corresponding atoms contained in the structural coordinates on top of one another show the same three-dimensional structure.

Moreover, the three-dimensional structural coordinates of GK protein in accordance with Table 1 and/or Table 2 are important information for drug design, and stored in a storage medium that can be read by computer in accordance with requirements, this information is processed with computer, and drug design is carried out. Accordingly, in another embodiment of this invention, a computer readable recording medium which recorded a program to function a computer as three dimensional coordinate memory means that memorises three-dimensional coordinates of amino acid residue in accordance with Table 1 and/or 2 is put forward.

Moreover, according to another embodiment of this invention, a computer readable recording medium that recorded a program which functions using computer as three dimensional coordinates memory means that memorised the three-dimensional coordinates of amino acid residue in accordance with Table 1 and/or 2, as binding site deduction means that deduces compound binding site of a protein having amino acid sequence represented by Sequence Number 8 and/or Sequence Number 5 using three dimensional coordinates of amino acid residue in accordance with Table 1 and/or 2 memorised in aforesaid three-dimensional coordinates memory measure, as binding compound memory means which memorised information about the type of compounds that bind to the protein and three-dimensional structure of aforesaid compounds, and as binding compound candidate selection means for selecting candidate compounds which are compatible to the compound binding site of the protein having amino acid sequence represented by Sequence Number 1 at least using the information about the three-dimensional structure of compound binding site of protein containing amino acid sequence represented by Sequence Number 8 and/or inferred Sequence Number 5 deduced by aforesaid binding site deduction means and the

information about three-dimensional structure of compound memorised in aforesaid binding compounds memory means, is put forward. Moreover, according to another embodiment of this invention, a computer equipped with aforesaid each means, is also put forward.

(Crystal of complex of GK protein with compound that binds to this).

Next, according to another embodiment of this invention, a crystal containing a complex of protein including amino acid sequence in accordance with Sequence Number 5 or Sequence Number 8 or amino acid sequence which is substantially the same amino acid sequence thereof with the compound which can bind to the said protein and a process for the production thereof are put forward.

When a compound which binds to GK protein is obtained, firstly, the GK protein and the compound thereof are mixed for example in an aqueous solution, and a complex is formed. As for the crystal of such complex, well known processes for the production of co-crystals such as co-crystallisation, soaking method or the like are used. As for the crystallisation condition and crystallisation process, refer to aforesaid processes.

For example, a compound that binds to GK protein is selected from the compound group represented by aforesaid formula (I).

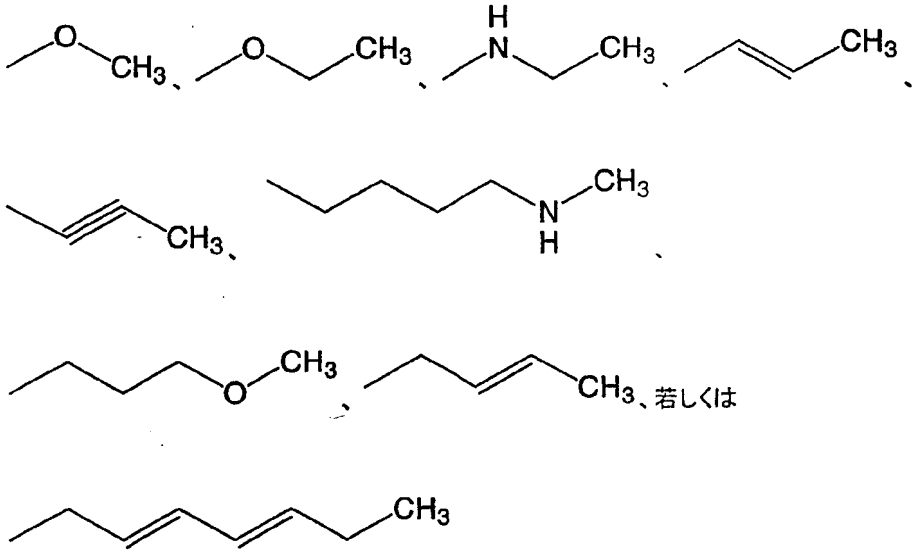
Wherein, as halogen atom of aforesaid formula (I), fluorine atom, chlorine atom, bromine atom, iodine atom or the like are exemplified, and among these, chlorine atom is preferred.

Moreover, as far as substituents in heteroaryl group of A, B of aforesaid formula (1) and formula (II) are concerned, amino group, carbamoyl group, carbamoyl amino group, carbamoyloxy group, carboxyl group, cyano group, sulphamoyl group, trifluoromethyl group, halogen atom, hydroxy group, formyl group, straight chained C1-C6 alkyl group, cyclic C3-C6 hydrocarbon group, aralkyl group, N-aralkyl amino group, N,N-diaralkyl amino group, aralkyloxy group, aralkyl carbonyl group, N-aralkyl carbamoyl group, aryl group, arylthio group, N-arylamino group, aryloxy group, aryl sulphonyl group, aryl sulphonyloxy group, N-arylsulfonylamino group, aryl sulphamoyl group, N-aryl carbamoyl group, aroyl group, aroxy group, C2-C6 alkanoyl group, N-C2-C6 alkanoyl amino group, C1-C6 alkylthio group, N-C1-C6 alkyl sulphamoyl group, N,N-di-C1-C6 alkyl sulphamoyl group, C1-C6 alkyl sulfinyl group, C1-

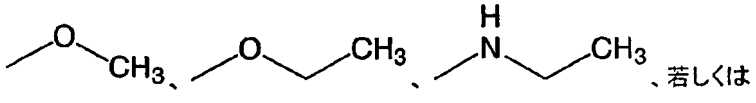
C6 alkylsulfonyl group, N-C1-C6 alkylsulfonyl amino group, C1-C6 alkoxy group, C1-C6 alkoxycarbonyl group or C1-C6 alkylamino group are denoted), or the like is nominated. Wherein, as for the preferably used substituent, amino group, carbamoyl group, carbamoyl amino group, carbamoyloxy group, carboxyl group, cyano group, sulphamoyl group, trifluoromethyl group, halogen atom, hydroxy group, formyl group, straight chained C1-C6 alkyl group or the like are exemplified.

Wherein, "hydrocarbon group" denotes 1-6 C straight chained alkyl group, or, among carbon atom constituting said alkyl group, a group in which 1 or 2, preferably 1 carbon atom may be substituted with nitrogen atom, sulfur atom or oxygen atom and/or carbon atom themselves in the said 1-6 C straight chain alkyl group may be bonded with double bond or triple bond. Number of said double bond or triple bond is preferably 1 or 2 and 1 is more preferred.

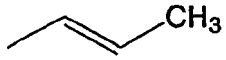
As said hydrocarbon group, in an embodiment, it is preferred to be methyl group, ethyl group, propyl group or isopropyl group, butyl group or a group represented by following formulae



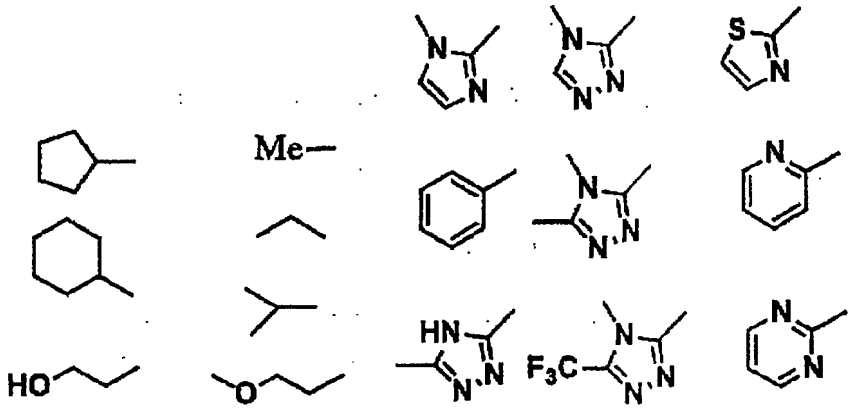
More preferred hydrocarbon group is methyl group, ethyl group, propyl group, isopropyl group or a group represented by following formulae



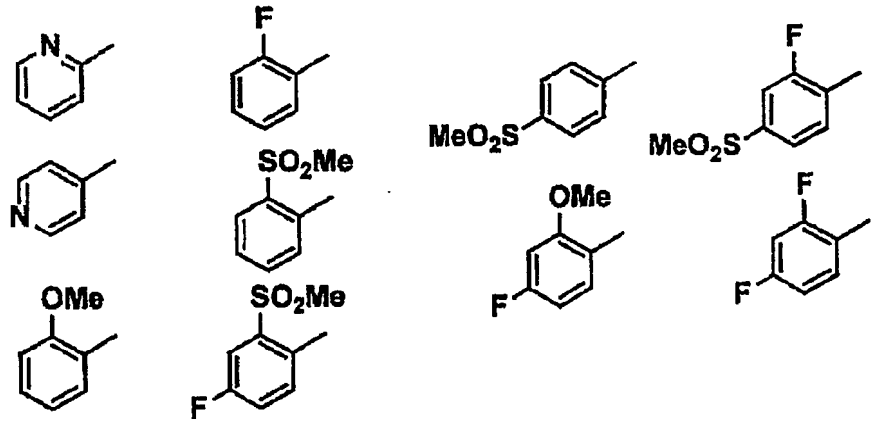
or



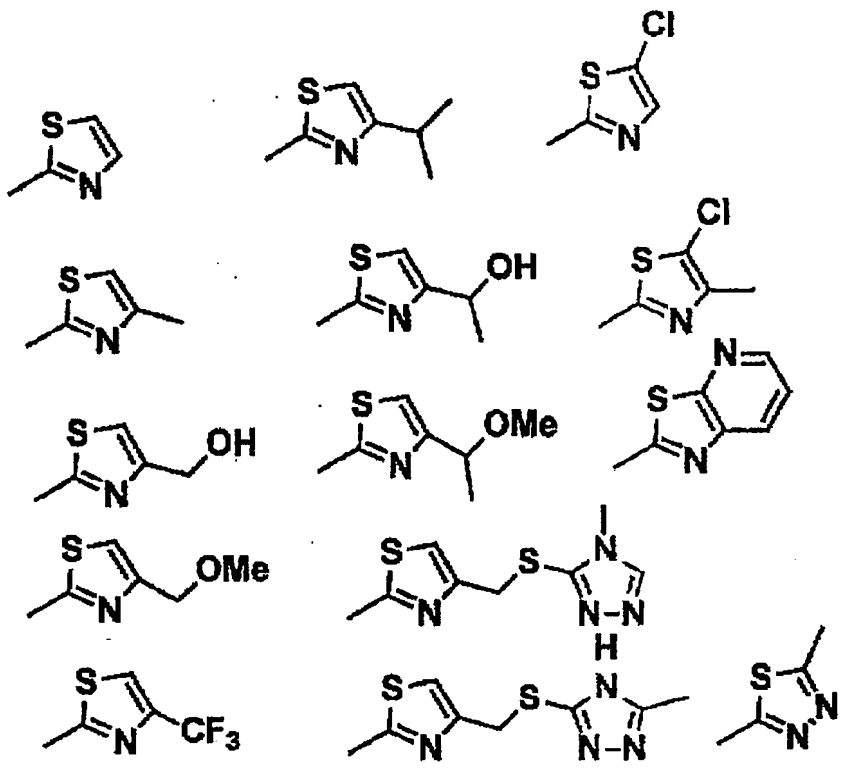
For example, as preferred A (in case of p = 0), the following groups are nominated.



As preferred B, for example, the following groups are nominated.



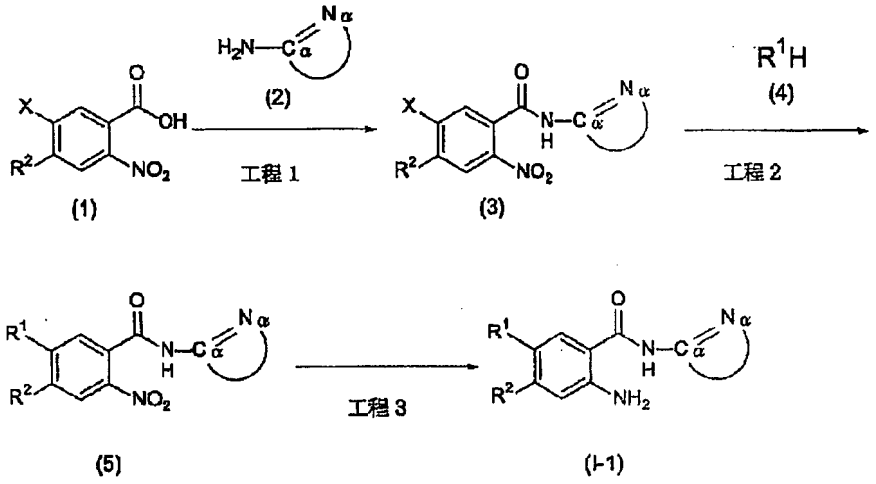
As heteroaryl group represented by formula (II), for example following heterocyclic groups are nominated.



Moreover, particularly preferred compounds are any of the compound represented by aforesaid formulae (IIIa)-(IIIc).

The compound of this invention (1) can be readily produced by using well known reaction means or according to well known method. Moreover, the compound of general formula (I) of this invention can be produced not only by synthesis in ordinary liquid phase, but also by synthesis using solid phase developed remarkably in recent years such as combinatorial synthesis method, parallel synthesis method or the like. Preferably, it can be produced for example using the following process.

Step 1 to Step 2 to Step 3



(wherein, each symbol is the same as in the aforesaid definition)

Step 1

This step is a process to produce compound (3) by reacting carboxylic acid compound (1) or reactive derivative thereof and amino compound containing optionally substituted monocyclic or bicyclic heteroaryl group represented by aforesaid formula (2) or salts thereof. In this reaction, ordinary amide formation reaction may be carried out by a method described in literature (for example Base and experiment of peptide synthesis, Shinya Izumiya et al., Maruzen, 1983, Comprehensive Organic Synthesis, vol 6, Pergamon Press Co. 1991 and the like) or in accordance with these, or by combining these and conventional method, in other word, it can be carried out by using condensing agent well-known for a person skilled in the art or by ester activation method, mixed acid anhydride method, acid chloride method, carbodiimide method and the like which can be used by a person skilled in the art. As such amide forming reagent, for example thionyl chloride, N,N-dicyclohexylcarbodiimide, 1-methyl-2-bromo pyridinium iodide, N,N'-carbonyldiimidazole, diphenyl phosphoryl chloride, diphenyl phosphoryl azide, N,N'-disuccinimidyl carbonate, N,N'-disuccinimidyl oxalato, 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride, ethylchloroformate, chloro formic acid isobutyl ester or benzo triazol-1-yl-oxy-tris (dimethylamino) phosphonium hexafluoro phosphate and the like are proposed, and wherein, for example thionyl chloride, N,N-dicyclohexylcarbodiimide or benzo triazol-1-yl-oxy-tris (dimethylamino) phosphonium hexafluoro phosphate and the like are suitable. Moreover, in amide forming reaction, a base, a condensation assistant may be used with the aforesaid amide forming reagent.

As base used, for example tertiary aliphatic amine such as trimethylamine, triethylamine, N,N-diisopropyl ethylamine, N-methylmorpholine, N-methylpyrrolidine, N-methylpiperidine, N,N-dimethylaniline, 1,8-diazabicyclo[5.4.0] undec-7-ene (DBU), 1,5-azabicyclo[4.3.0] non-5-ene (DBN) or the like, for example aromatic amine such as pyridine, 4-dimethylaminopyridine, picoline, lutidine, quinoline, isoquinoline and the like are proposed, and wherein, for example tertiary aliphatic amine and the like is preferred, and in particular, for example triethylamine or N,N-diisopropyl ethylamine and the like is suitable.

As condensation assistant used, for example N-hydroxybenzotriazole hydrate, N-hydroxy succinimide, N-hydroxy-5-norbornene-2,3-dicarboximide or 3-hydroxy-3,4-dihydro-4-oxo-1,2,3-benzotriazole and the like are proposed, and among these, for example N-hydroxybenzotriazole and the like are suitable.

The amount of amino compound (2) used differs depending on the kind of compound and solvent used and other reaction conditions, however, usually, 0.02 to 50 equivalents, preferably 0.2 to 2 equivalents with respect to 1 equivalent of carboxylic acid compound (1) or reactive derivative thereof. Herein, as reactive derivative, for example active ester derivative, active amide derivative and the like which are used in the sphere of usual organic chemistry are nominated.

The amount of used amide forming reagent differs depending on the kind of compound and solvent used and other reaction conditions, however, usually 1-50 equivalents, preferably 1-5 equivalents with respect to 1 equivalent of carboxylic acid compound (1) or reactive derivative thereof.

The amount of used condensation assistant differs depending on the kind of compound and solvent used and other reaction conditions, however, usually it is 1-50 equivalents, preferably 1-5 equivalents with respect to 1 equivalent of carboxylic acid compound (1) or reactive derivative thereof.

The amount of used base differs depending on the kind of compound and solvent used and other reaction conditions, however, usually 1 to 50 equivalents, preferably 3 to 5 equivalents.

The reaction solvent used in this step, is for example insert organic solvent, and it is not restricted in

particular so long as it does not hinder the reaction. However, in an embodiment, for example methylene chloride, chloroform, 1,2-dichloroethane, trichloroethane, N,N-dimethylformamide, acetic acid ethylester, acetic acid methylester, acetonitrile, benzene, xylene, toluene, 1,4-dioxane, tetrahydrofuran, dimethoxyethane or a mixed solvent thereof are proposed, however, in particular for example methylene chloride, chloroform, 1,2-dichloroethane, acetonitrile, N,N-dimethylformamide or the like are suitable in term of securing a suitable reaction temperature.

The reaction temperature is -100°C to boiling point of solvent, preferably 0 to 30°C.

The reaction time is 0.5 to 96 hours, preferably 3 to 24 hours.

The base, amide formation reagent, condensation assistant used in this step 1 can be used as a single species or in combination of two or more.

When the compound (3) contains protecting group, said protecting group can be suitably eliminated. Elimination of aforesaid protecting group can be carried out by method described in literature (Protective Groups in Organic Synthesis, written by T.W. Green, the second edition, John Wiley & Sons Co, 1991, or the like) or method in accordance with this or by combining these and conventional method.

Compound (3) obtained in this way can be provided for the next step by isolating and purifying with well known separation and refinement means, for example concentration, vacuum concentration, crystallisation, solvent extraction, re-precipitation, chromatography and the like or without isolating and purifying.

Step 2

This step comprises a process to produce compound (5) by reacting amide compound (3) obtained in aforesaid step 1 and compound (4).

In this reaction, a base may be added to the reaction system in accordance with requirements. As used compound (4), preferably phenol derivative or thiol derivative is preferred. As said phenol derivative or thiol derivative, for example phenol, thiophenol, thio imidazole, thio triazole and the like are

nominated. The amount of compound (4) used differs depending on the kind of compound and solvent used or other reaction conditions, however, usually it is 2-50 equivalents, preferably 2-5 equivalents with respect to 1 equivalent of amino derivative (3). As used base, for example tertiary aliphatic amine such as trimethylamine, triethylamine, N,N-diisopropyl ethylamine, N-methylmorpholine, N-methylpyrrolidine, N-methylpiperidine, N,N-dimethylaniline, 1,8-diazabicyclo[5.4.0] undec-7-ene (DBU), 1,5-azabicyclo[4.3.0] non-5-ene (DBN) or the like, for example aromatic amine such as pyridine, 4-dimethylaminopyridine, picoline, lutidine, quinoline, isoquinoline and the like, alkali metal such as metallic potassium, metallic sodium, metallic lithium and the like, alkali metal hydride such as sodium hydride, potassium hydride and the like, alkali metal alkylate such as butyl lithium and the like, alkali metal alkoxide such as potassium-tert-butyrate, sodium ethylate or sodium methylate and the like, alkali metal hydroxide such as potassium hydroxide, sodium hydroxide and the like, alkali metal carbonate such as potassium carbonate and the like are nominated, among these for example tertiary aliphatic amine, alkali metal hydride or alkali metal carbonate are preferred, and in particular, for example triethylamine, N,N-diisopropyl ethylamine, sodium hydride or potassium carbonate are suitable.

The amount of aforesaid base used differs depending on the kind of compound and solvent used and other reaction conditions, however, it is usually 0 to 50 equivalents, preferably 2-10 equivalents with respect to 1 equivalent of amide compound (3). Said base can be used as a single species or two or more species in accordance with requirements.

As used insert organic solvent, there are no restrictions in particular so long as the reaction is not hindered. However, in an embodiment, for example methylene chloride, chloroform, 1,2-dichloroethane, trichloroethane, N,N-dimethylformamide, N,N-dimethyl acetamide, acetic acid ethylester, acetic acid methylester, acetonitrile, benzene, xylene, water, toluene, 1,4-dioxane, tetrahydrofuran, or a mixed solvent thereof are proposed.

Compound (5) obtained in this way can be and isolated and purified with well known separation and refinement means, for example concentration, vacuum concentration, crystallisation, solvent extraction, re-precipitation, chromatography and the like.

Step 3

This step is a process to produce compound (I) used in this invention by reduction of compound (5). As for reductive reaction used in this step, well-known processes to a person skilled in the art are used. As the reductive reaction used in this step, in an embodiment, for example (1) catalytic reduction method using hydrogen, formic acid, ammonium formate, hydrazine hydrate and palladium, platinum, nickel catalyst, (2) reduction method using hydrochloric acid, ammonium chloride and iron, (3) reduction method using methanol and tin chloride are nominated.

The amount of reducing agent used in the aforesaid reductive reaction differs depending on the kind of compound and solvent to be used and other reaction conditions, however, it is usually 1-50 equivalents, preferably 2-20 equivalents with respect to 1 equivalent of compound (5).

The reaction solvent used is not restricted in particular so long as the reaction is not hindered. However, for example halogenated hydrocarbons such as dichloromethane, chloroform and the like, ethers such as diethyl ether, tert-butyl methyl ether, tetrahydrofuran and the like, amides such as N,N-dimethylformamide, N,N-dimethylacetamide and the like, sulphoxides such as dimethylsulfoxide and the like, nitriles such as acetonitrile and the like, an alcohol such as methanol, ethanol, propanol and the like, aromatic hydrocarbons such as benzene, toluene, xylene and the like, water or mixed solvent thereof can be used.

Reaction temperature and the reaction time are not restricted in particular. However, the reaction is carried out for 1-20 hours approx. and preferably 1 to 5 hours approx. at a reaction temperature of -10 to 100°C approx. and preferably 0 to 50°C approx.

Compound (1) used in this invention obtained in this way can be provided for the next step by isolating and purifying with well known separation and refinement means, for example concentration, vacuum concentration, crystallisation, solvent extraction, re-precipitation, chromatography and the like or without isolating and purifying.

Compound of aforesaid each step may contain protecting group on each substituent. Aforesaid protecting group can be suitably eliminated in each step using well known method, method in accordance with that or method combined these and the conventional method. As for the embodiment of

climination, suitablc climination rcaction is possiblc dcpcnding on thc kind of compound, rcaction and other reaction conditions. However, it is considered the case in which each protecting group is eliminated individually and the case in which each protecting group is simultaneously eliminated and the like, and it can be suitably selected by a person skilled in the art. As aforesaid protecting group, for example protecting group of hydroxy group, protecting group of amino group, protecting group of carboxyl group, protecting group of aldehyde, protecting group of keto group and the like are nominated. Moreover, the order of elimination aforesaid protecting groups is not limited in particular.

As protecting group of hydroxy group, for example lower alkyl silyl group such as tert-butyldimethylsilyl group, tert-butyl diphenyl silyl group and the like, for example lower alkoxymethyl group such as methoxy methyl group, 2-methoxyethoxymethyl group and the like, for example aralkyl group such as benzyl group, p-methoxybenzyl group and the like, for example acyl group such as formyl group, acetyl group and the like are proposed, and among these, tert-butyldimethylsilyl group, acetyl group and the like are in particular preferred.

As protecting group of amino group, for example aralkyl group such as benzyl group, p-nitrobenzyl and the like, for example acyl group such as formyl group, acetyl group and the like, for example lower alkoxycarbonyl group such as ethoxycarbonyl group, tert-butoxycarbonyl group and the like, for example aralkyloxy carbonyl group such as benzyloxycarbonyl group, p-nitrobenzyl oxycarbonyl group and thc likc arc proposcd, and among thcsc, nitrobcnzyl group, tcrt-butoxy carbonyl group, benzyloxycarbonyl group and the like are particularly preferred.

As protecting group of carboxyl group, for example lower alkyl group such as methyl group, ethyl group, tert-butyl group and the like, for example aralkyl group such as benzyl group, p-methoxybenzyl group and the like are nominated, and among these, methyl group, ethyl group, tert-butyl group, benzyl group and the like are particularly preferred.

As protecting group of keto group, for example dimethyl ketal group, 1,3-dioxirane group, 1,3-dioxolane group, 1,3-dithiane group, 1,3-dithiorane group and the like are proposed, and among these, dimethyl ketal group, 1,3-dioxolane group and the like are more preferred.

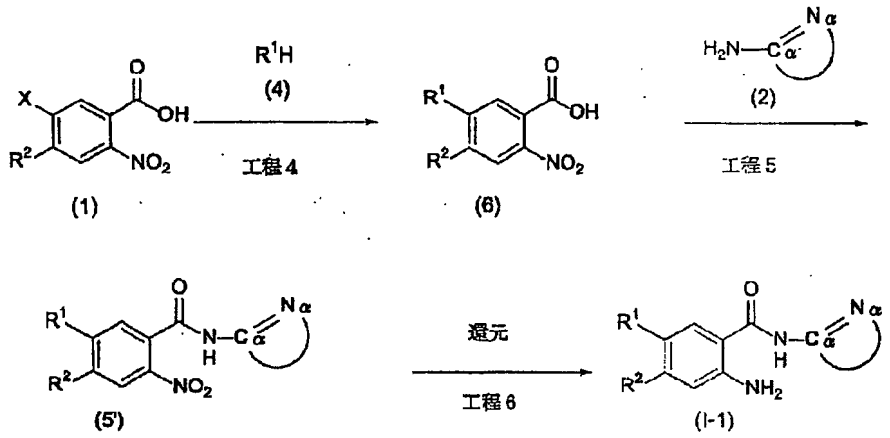
As protecting group of aldehyde group, for example, dimethylacetal group, 1,3-dioxirane group, 1,3-dioxolane group, 1,3-dithiane group, 1,3-dithiorane group and the like are proposed, and among these, dimethylacetal group, 1,3-dioxolane group and the like are more preferred.

In the production of compound used in this invention, there is a case that the protecting group is introduced to functional group in order to proceed reaction with good efficiency. The introduction of these protecting groups can be suitably selected by a person skilled in the art, and elimination of aforesaid protecting groups can be carried out by a method described in aforesaid Protective Groups In Organic Synthesis and the like, a method in accordance with that or by combining that and conventional method. Moreover, the order of elimination of protecting groups can be suitably selected by a person skilled in the art.

Compound (1) obtained in this way can be subjected to the next step after isolating and purifying with well known separation and refinement means, for example concentration, vacuum concentration, crystallisation, solvent extraction, re-precipitation, chromatography and the like or without isolating and purifying.

Moreover, the compound used in this invention (I) can be also produced by the following step.

Step 4 to Step 5 to Step 6



(wherein, each symbol has same aforesaid definition).

As far as the aforesaid Step 4, Step 5 and Step 6 are concerned, it can be carried out using the same amount of reagent, reaction solvent, reaction temperature and other reaction conditions as in aforesaid Step 1, Step 2 and Step 3.

When a protecting group is necessary for R2, it can be carried out by a person skilled in the art suitably selecting a process from a method described in aforesaid Protective Groups In Organic Synthesis and the like, a method in accordance with that or by combining that and conventional method.

Compounds (6) and (5') obtained in this way can be provided for the next step after isolating and purifying with well known separation and refinement means, for example concentration, vacuum concentration, crystallisation, re-precipitation, solvent extraction and the like or without isolating and purifying.

Compound (1) used in this invention can be isolated and purified by well known separation and refinement means, for example concentration, vacuum concentration, crystallisation, re-precipitation, solvent extraction and the like.

In aforesaid step 1 to 6, the elimination of protecting groups differ depending on the kind of aforesaid protecting group and stability of compound, however, it can be carried out by aforesaid method described in Protective Groups in Organic Synthesis, written by T.W. Green, the second edition, John Wiley & Sons Co, 1991, or the like or a method in accordance with this or by combining these and conventional method. For example, it can be carried out by solvolysis using acid or base, chemical reduction using hydrogenated metallic complex and the like or catalytic reductions using palladium carbon catalyst, Raney nickel and the like.

The benzamide compound put forward by this invention can exist as pharmacologically acceptable salt. Aforesaid salt can be produced in accordance with conventional methods. In an embodiment, when aforesaid compound (1) contains basic group derived from for example amino group, pyridyl group within the molecule, it can be converted to corresponding pharmacologically acceptable salt by treating aforesaid compound with an acid.

As aforesaid acid addition salt, for example the acid addition salt of halide acid salt such as hydrochloride, hydrofluoric acid salt, hydrobromic acid salt, hydroiodic acid salt or the like, inorganic acid salt such as nitrate, perchlorate, sulfate, phosphate, carbonate or the like, lower alkyl sulfonate such as methanesulfonate, trifluoromethanesulfonate, ethanesulfonic acid salt or the like, aryl sulfonate such as benzensuplhonate, p-toluenesulfonate or the like, organic salt such as fumarate, succinate, citrate, tartrate, oxalate, maleate or the like and organic acid of amino acid or the like such as glutamic acid salt, aspartate or the like are nominated. Moreover, when the compound of this invention is containing acidic group in aforesaid group, for example when containing carboxyl groups, it can be converted to corresponding pharmacologically acceptable salt by treating aforesaid compound with a base. As aforesaid base addition salt, for example salt of alkali metal salt such as sodium, potassium and the like, alkaline earth metal salt such as calcium, magnesium and the like, organic base such as ammonium salt, guanidine, triethylamine, dicyclohexylamine and the like are nominated. Furthermore, the compound of this invention may exist as free compound or arbitrary hydrate or solventate of salts thereof.

In accordance with this invention, as explained in detail in the description of Examples, crystal of complex of GK protein containing amino acid sequence shown in Sequence Number 5 and compounds of aforesaid formula (IIIa)-formula (IIIc) are obtained. In GK protein shown in Sequence Number 5, it has been elucidated that compound binding site is constituted from the amino acid residue of tyrosine 61 - serine 69, glutamic acid 96 - glutamine 98, isoleucine 159, methionine 210 - tyrosine 215, histidine 218 - glutamic acid 221, methionine 235, arginine 250, leucine 451 - lysine 459 by analysing these crystal three - dimensional structure coordinates.

Moreover, according to another embodiment of this invention, a process for the production of crystal containing a complex of protein and a compound that binds to the protein, wherein a protein production step to produce a protein containing amino acids sequence in which prescribed number of amino acid residues at N terminal side and/or C terminal side are deleted as described above from the protein containing amino acid sequence in accordance with Sequence Number 2 and a step to cause reaction of a compound which binds to the protein obtained in aforesaid protein production step with the protein obtained in said protein production step are included, is put forward.

As protein to be produced in the aforesaid protein production step, the number thereof is not restricted as long as it is within a range that steric hindrance between adjacent GK proteins in the crystal is eliminated. In an embodiment, for example, in the amino acids sequence shown in Sequence Number 2, the amino acids sequences in which amino acid residues at N terminal side are deleted in numbers of 1-50, preferably 3-30, more preferably 5-25, still more preferably 8-18, even more preferably 11-15 or the like are nominated. Moreover, the amino acid sequences in which amino acid residues at C terminal side of 1-8, preferably 1-7, more preferably 2-6 or the like are deleted, are nominated.

(The drug design process using three-dimensional structural coordinate).

Three-dimensional structure of GK protein of this invention obtained as above provides important information for drug creation system using CARDD (Computer Aided Rational Drug Design). It is an important step of the target drug creation and development to elucidate the active center and allosteric site of this GK protein and to search for a compound which is compatible to said site, interacts with the GK protein and thereby activates to inhibits the GK protein.

In other words, according to another embodiment of this invention, a drug design process of the kind to design structure of compound that binds to said protein based on stereostructure information of protein, characterised in that the stereostructure information of said protein comprises information to be obtained by analysing the crystal obtained as described above, is put forward. As such drug design process, there are techniques to make drug design using energy calculation, activity prediction value analogous to this or pharmacophore and a technique to visually design drug using computer graphics technique.

As process by technique using energy calculation, activity prediction value analogous to this or pharmacophore, (1) a drug design process including a binding site deduction step to deduce compound binding site of said protein based on stereostructural information obtained as above and a selection step to select a compound compatible to the compound binding site deduced in aforesaid binding site deduction step from the compound library, (2) a drug design process including a binding site deduction step to deduce compound binding site of said protein based on aforesaid stereostructural information and a compound structure assembly step to construct a structure of compound compatible to the compound binding site deduced in aforesaid binding site deduction step, or the like are exemplified.

As process to deduce compound binding site of aforesaid protein, for example, a process wherein the ligand bonded site in the co-crystal of compound is identified by confirming with visual observation on display of computer, and in addition to that, a process wherein the site to which ligand is likely to bind is identified with respect to the protein crystal structure solved under the condition that ligand is not bound, are nominated. In any processes, well-known method and commercial computer soft wear can be used. In former process, for example, it is possible to use software such as Insight II (Accelrys Inc.), SYBYL (Tripos Inc.), MOE (Chemical Computing Group) or the like. On the other hand, For example, in latter process, well known technique such as Cavity search: an algorithm for the isolation and display of cavity-like binding regions, (Journal of Computer-Aided Molecular Design. 4(4): 337-54, 1990) or the like can be used, and it can be carried out using software such as SiteID (Tripos Inc.) or the like.

Once the binding site of compound in protein was able to be deduced, a compound which can be compatible to the deduced binding site is selected. As process to select this candidate compound, structural information of the compound is acquired from existing compound library, and bindable candidate compound is selected by comparing the structural information of compound in the library and structural information of the binding site deduced as above.

In a further embodiment, 1 or more residues of amino acid residues of amino acid sequence shown in Sequence Number 5 (tyrosine 61 - serine 69, glutamic acid 96 - glutamine 98, isoleucine 159, methionine 210 - tyrosine 215, histidine 218 - glutamic acid 221, methionine 235, arginine 250, leucine 451 - lysine 459) or pharmacophore of hydrogen bonding or hydrophobic properties or the like formed from the functional group of ligand in the complex, and also the protein surface produced from the protein structure or a structure in which the orientation of a part of the side chain thereof is modified, are used as search condition, and the conformation and orientation of each compound is systematically searched from the compound library, and whether the conditions are satisfied or not is judged and it is selected.

As an alternative process, while systematically searching the conformation and orientation of each compound from the compound library, the candidate compound is virtually docked with respect to the structure of ligand binding site constructed from the amino acid residues (tyrosine 61 - serine 69, glutamic acid 96 - glutamine 98, isoleucine 159, methionine 210 - tyrosine 215, histidine 218 - glutamic acid 221, methionine 235, arginine 250, leucine 451 - lysine 459) or a structure in which the orientation

of a part of the side chain thereof is modified, the species that formed interaction of close proximity of 4 Å or less with 1 or more residues of amino acid residues of amino acid sequence (tyrosine 61 - serine 69, glutamic acid 96 - glutamine 98, isoleucine 159, methionine 210 - tyrosine 215, histidine 218 - glutamic acid 221, methionine 235, arginine 250, leucine 451 - lysine 459) is selected, or selection is carried out using energy evaluation function.

On the other hand, the candidate compound can also be selected by designing a bindable compound based on the structural information of the binding site deduced as above. In a further embodiment, each atomic species and functional groups are connected so that interactions are possible with respect to the structure of ligand binding site constructed from the amino acid residues (tyrosine 61 - serine 69, glutamic acid 96 - glutamine 98, isoleucine 159, methionine 210 - tyrosine 215, histidine 218 - glutamic acid 221, methionine 235, arginine 250, leucine 451 - lysine 459) or a structure in which the orientation of a part of the side chain thereof is modified, and thereby a chemical structure is constructed. As this process, a process wherein chemical groups such as methyl, ethyl and the like are arranged in the active site and a compatible compound is searched, and a process wherein atoms are bonded at active site using a computer program.

Moreover, with the process by energy evaluation using computer, for example a process to determine the bond energy of a compound and GK protein using molecular force field calculation is nominated. The calculation thereof is applied to each compound in database, and candidate compounds which can form stable binding are selected from the library compound. With some computer programs, such as Ludi of Insight II, when three-dimensional structural coordinates of interacting amino acid residues in the protein molecule are input, candidate of bindable compounds are automatically selected and output, and it can be suitably used.

Moreover, as far as the drug design on the basis of three-dimensional structure of molecule is concerned, many literature are known including development of pharmaceutical Vol. 7 "molecular design" (Hirokawa Shoten). In an embodiment, first, using flexible ligand binding simulation software such as for example FlexiDock, FlexX or the like, a library of low molecular (molecular weight 1000 or less) compounds (for example about 150000 species) can be screened with computer. For chemical compounds in this library, three-dimensional structure is built using a program such as CONCORD or the

like, and compounds compatible to the active site can be selected.

On the other hand, as a process of visual drug design, a drug design process characterised in including a binding site deduction step to deduce compound binding site of said protein based on aforesaid stereostructural information and a design step wherein the structure of the compound is visually designed so that aforesaid compound binding site deduced in aforesaid binding site deduction step and a compound compatible to said compound binding site can interact, is nominated. For example, structure assembly or structure modification is carried out with respect to the structure of ligand binding site constructed from the amino acid residues (tyrosine 61 - serine 69, glutamic acid 96 - glutamine 98, isoleucine 159, methionine 210 - tyrosine 215, histidine 218 - glutamic acid 221, methionine 235, arginine 250, leucine 451 - lysine 459) or a structure in which the orientation of a part of the side chain thereof is modified, so that it can interact with 1 or 2 or more residues among these residues.

In an embodiment, with visual process, first, crystal structure of the complex of GK protein and a compound bound to this is displayed on a computer screen according to the obtained structural coordinates. And, while considering the chemical interaction, the binding possibilities of the compounds in the library and GK protein are successively examined on computer. Wherein, the chemical interactions to be considered are electrostatic interaction, hydrophobic interaction, hydrogen bonding, van der Waals interaction or the like. In other words, the structure in three dimensional space of said compound is generally considered whether a structure favourable for the interaction is formed or not, so that among the functional groups thereof, the groups likely to be negatively charged such as carboxyl group, nitro group, halogen group or the like interact with amino acid residues having positive charge such as lysine, arginine, histidine of GK protein, the groups likely to be positively charged such as amino group, imino group, guanidyl group or the like interact with amino acid residues having negative positive charge such as glutamic acid, aspartic acid of GK protein, hydrophobic functional groups such as aliphatic group and aromatic group interact with hydrophobic amino acid residues such as alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine, the groups which participate in hydrogen bond such as hydroxy group, amide group or the like can for hydrogen bonding with the main chain or side chain part of the GK protein, furthermore steric hindrance is not caused by the binding of said compound and GK protein, moreover, furthermore, the void part is filled so that the void part is eliminated as much as possible so that the van der Waals interaction is increased, or the like. In this way,

the factors such as electrostatic interaction, hydrophobic interaction, van der Waals interaction, hydrogen bond or the like are comprehensively considered visually on the computer screen, and finally, whether the candidate compound can bind to the GK protein or not is determined.

As program for selecting compound candidate by visual observation in this way, simulation programs such as Insight II and MOE or the like are exemplified. In order to generate promising candidate compounds that interact with GK protein, the candidate compounds are contacted with GK protein, and activity of GK protein is measured. In practice, the promising candidate compound is mixed with GK protein, crystallised, and whether it is compatible or not is evaluated. Further, more desirable structure is formed by modifying the compatible complex using organic synthesis.

Moreover, the visual technique and the technique that considers energy may be suitably combined, and used. As such computer software, flexiDock (Tripos Inc.), FlexX (Tripos Inc.), SYBYL (Tripos Inc.), Insight II (Accelrys Inc.), MOE (Chemical Computing Group Inc.) or the like are nominated.

Moreover, in accordance with this invention, the compounds selected by aforesaid drug design process are synthesized, and these compound groups can be provided as compound array (compound library). Because a large quantity of candidate compounds can be assayed at one time using a technique such as high through-put screening or the like, the inhibitor or activator of glucokinase can be screened with good efficiency.

(Compounds obtained by process of this invention and therapeutic agent including these)

The compounds designed by aforesaid drug design process have has ability to bind to glucokinase, therefore they can be used as activators of glucokinase or glucokinase inhibitors. Moreover, the therapeutic agent or medicinal composition containing such compound can be effectively used as therapeutic agent of disease involving glucokinase activity (for example diabetes mellitus therapeutic agent).

Aforesaid medicinal composition contains a compound that binds to glucokinase of this invention as effective ingredient in pharmacologically effective dose thereof together with suitable pharmacologically permitted support or diluent. As the pharmacologically acceptable support which can be used in aforesaid

medicinal composition (drug formulation), diluent such as filler, extender, binding agent, humectant, disintegrating agent, surface active agent, lubricant or the like which are conventionally used corresponding to the form of the formulation or excipient or the like are exemplified. These carriers can be suitably selected and used corresponding to administration unit form of the obtained formulation.

As administration unit form of medicinal composition of this invention, various forms can be selected according to therapeutic purpose, and, as representative examples thereof, solid administrative form such as tablet, pill, powder, powder agent, granule, encapsulated formulation or the like and liquid agent administrative form such as solution, suspending agent, emulsion, syrup, elixir or the like are included. Further these are classified into orally administered agent, aoral drug, transnasal agent, vaginal agent, suppository, sublingual agent, ointment or the like according to administration route, and it can be formulated, molded and prepared each according to conventional process. For example, when it is formed to a tablet form, excipient such as lactose, lactose, refined sugar, sodium chloride, glucose, urea, starch, calcium carbonate, kaolin, crystalline cellulose, silicic acid, potassium phosphate or the like, binding agent such as water, ethanol, propanol, simple syrup, glucose syrup, starch solution, gelatin solution, carboxymethylcellulose, hydroxypropylcellulose, methylcellulose, polyvinylpyrrolidone or the like, disintegrating agent such as carboxymethylcellulose sodium, carboxymethylcellulose calcium, low degree of substitution hydroxypropylcellulose, dried starch, sodium alginate, agar powder, laminaran powder, sodium bicarbonate, calcium carbonate or the like, surface active agent such as polyoxyethylene sorbitan fatty acid ester species, lauryl sodium sulfate, stearic acid monoglyceride or the like, disintegration inhibitor such as refined sugar, stearin, cacao butter, hydrogenation oil or the like, absorption accelerating agent such as quaternary ammonium base, sodium lauryl sulfate or the like, humectant such as glycerol, starch or the like, adsorbent such as starch, lactose, kaolin, bentonite, colloidal silica or the like, lubricant or the like such as purified talc, stearate, boric acid powder, polyethyleneglycol or the like can be used. Further, the tablet can be formed into a tablet coated with ordinary agent coating in accordance with requirements, for example sugar coated tablet, gelatin encapsulation tablet, enteric coated tablet, film coatings tablet and moreover can be made into double layer tablet or multilayer tablet.

When a form of pill is formed, as formulation carrier, for example, excipient such as glucose, lactose, starch, cacao butter, hardened vegetable oil, kaolin, talc or the like, binding agent such as powdered gum

arabic, tragacanth powder, gelatin, ethanol or the like, disintegrating agent or the like such as laminaran, agar or the like can be used.

For encapsulated formulation, effective ingredient of this invention is mixed with the various formulation carrier exemplified as above according to normal method, and it is prepared by being packed into hard gelatin capsule, soft capsule or the like.

The liquid administration form for oral administration includes pharmacologically permitted solution, emulsion, suspension, syrup, elixir or the like containing generally used inert diluent, for example water, and furthermore, auxiliary such as wetting agent, emulsion, suspending agent or the like can be contained, and these are prepared according to normal method.

For the preparation of liquid administrative form for aoral administration, for example, sterile aqueous or non-aqueous solution, emulsion, suspension or the like, as diluent, for example water, ethanol, propylene glycol, polyethyleneglycol, ethoxylation isostearyl alcohol, polyoxyisosteary alcohol, polyoxyethylene sorbitan fatty acid ester and vegetable oil or the like such as olive oil or the like can be used, and moreover, injectable organic ester species, for example, ethyl oleate or the like can be formulated. Further, ordinary solubilser, buffer agent, wetting agent, emulsifier, suspending agent, preservative, dispersant or the like can be added to these. Sterilisation can be carried out for example by filtration operation through bacteria retaining filter, formulation of fungicide, irradiation treatment and heat treatment or the like. Moreover, these can be prepared as sterile solid composition which can be dissolved in sterile water or suitable sterilisable vehicle immediately before the use.

When forming into a form of suppository or vaginal administration, as formulation carrier, for example polyethyleneglycol, cacao butter, higher alcohol, higher alcohol ester, gelatin and semi-synthetic glyceride or the like can be used.

When forming into a form ointment such as paste, cream, gel or the like, as diluent, for example white petrolatum, paraffin, glycerol, cellulose derivative, propylene glycol, polyethyleneglycol, silicon, bentonite and vegetable oil or the like such as olive oil or the like can be used.

A composition for transnasal or sublingual administration can be prepared according to conventional method using standard excipient.

Moreover, in agent of this invention, colorant, preservative, flavor, flavor agent, sweetener or the like or other pharmaceutical or the like can be contained in accordance with requirements.

The amount of the effective ingredient to be contained in the aforesaid drug formulation and dose thereof are not restricted in particular, and it is suitably selected from a wide range corresponding to the desired therapy effect, administration method, therapy period, age, sex of patient, other conditions or the like. In general, the dose is about 0.01 mg - 1000 mg, preferably about 1 mg - 100 mg per 60 kg in weight per day usually, and it can be administered once or divided into several times per day.

Sequence number of sequence table of this specification shows following sequence.

(Sequence number: 1).
Base sequence of DNA encoding human derived liver type glucokinase is shown.

(Sequence number: 2).
Amino acid sequence of human derived liver type glucokinase is shown.

(Sequence number: 3).
Amino acid sequence of human derived beta cell glucokinase is shown.

(Sequence number: 4).
Base sequence of DNA encoding the protein in which 11 amino acid residues at N terminal side of human derived liver type glucokinase are deleted, is shown.

(Sequence number: 5).
Amino acid sequence of the protein in which 11 amino acid residues at N terminal side of human derived liver type glucokinase are deleted, is shown.

(Sequence number: 6).

Base sequence of primer-1 used in PCR reaction in the following Example 1 is shown.

(Sequence number: 7).

Base sequence of primer-2 used in PCR reaction in the following Example 1 is shown.

(Sequence number: 8).

Amino acid sequence of the protein in which 15 amino acid residues at N terminal side of human derived liver type glucokinase are deleted, is shown.

(Sequence number: 9).

Base sequence of the primer used in PCR reaction in the following Example 6 is shown.

(Sequence number: 10).

Base sequence of the primer used in PCR reaction in the following Example 6 is shown.

Examples

Hereinafter, this invention will be described in concrete terms using Examples.

A process for purification of mutant type enzyme

In human glucokinase, there are liver type and pancreas type depending on the difference of promoter, and 15 residues at N terminal are different. In order to carry out crystallisation for the purpose of three-dimensional structure analysis, a mutant type enzyme which lacked a part or a whole of this region was made by the following process.

PCR reaction was carried out using cDNA of human liver type glucokinase cloned on pCR2.1 (made by INTROGEN Co.) and two kinds of primer sets, comprising

a combination of 5'-gtcacaaggagccagaagcttatggccttgactctggtag-3' (sequence number 6) and
5'-gaagccccacgacattgttcccttctgc-3 (sequence number 7), and

a combination of 5'-ccaggcccagacagccaagcttatggtagagcagatcc-3' (sequence number 9) and
5'-gaagccccacgacattgttcccttctgc 3' (sequence number 10).

The Hind III, Cla I fragment of the obtained PCR product was substituted with Hind III-Cla I region of liver type GK cloned at Hind III, Eco RI site of pFLAG/CTC vector (Eastman Kodak), and thereby cDNAs encoding mutant type GK (Δ1-11) which lacked 1-11 residues of liver type GK and mutant type GK (Δ1-15) which lacked 1-15 residues were obtained. The sequence of the obtained DNA was confirmed, and thereafter, these vectors were made as expression vectors, and Escherichia coli DH alpha strain (made by Takara Shuzo company) was transformed.

Transformant was cultured in LB medium at 37°C till the absorption at 600 nm became 0.8, and thereafter, isopropyl-1-thio-beta-D-galactoside (made by Wako Pure Chemicals Co.) was added so as to become the final concentration of 0.4 mM, and the protein production was induced at 25°C for 16 hours.

Cultured Escherichia coli was collected using centrifuge, and it was suspended in a buffer containing the following components (50 mM potassium phosphate (Potassium Phosphate) pII 7.5, 50 mM NaCl, 2 mM DTT, 0.5 mM Pefabloc SC (made by Kanto Chemicals Company), a proteinase inhibitor mixture (made by Roche Co.)) .

Collected Escherichia coli was pulverised by ultrasonic pulverisation method, and soluble fraction was dialysed against aforesaid buffer, and thereafter, it was purified using HiTrapQ column (made by Amersham Corp.). The GK fraction eluted from HiTrapQ column by potassium chloride gradient was diluted to a salt concentration of 50 mM by dilution.

The diluted GK fraction was purified by Glucosamin Sepharose column produced by a process demonstrated in the paper (Preparative Biochemistry, 20(2), 163-178 (1990)). The GK fraction was adsorbed onto Glucosamin Sepharose column, and impurity was eliminated with 100 mM sodium chloride, and thereafter, it was eluted by glucose of 1 M.

The eluted GK fraction was refined by MonoQ10/10 column. The GK fraction eluted from the MonoQ10/10 column (made by Amersham Corp.) by sodium chloride gradient was purified by Superdex 200 column (made by Amersham Corp.) using 50 mM Tris-Cl pH 7.2, 50 mM NaCl buffer as mobile layer.

Crystallisation process

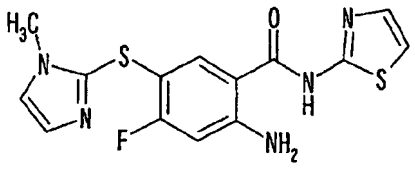
Crystal of mutant type GK [Δ1-11] / glucose / compound complex

The crystal of mutant type GK (Δ1-11) / glucose / compound complex was obtained using a technique of the following vapor diffusion. Moreover, mutant type GK (Δ1-11) denotes a glucokinase containing amino acid sequence represented by Sequence Number 5.

In other words, mutant type GK purified to a high purity was concentrated, and finally a solution of mutant type GK of about 10 mg/ml (25 mM Tris-Cl, 50 mM NaCl, 5 mM TCEP) was formed. Thereto were added glucose of final concentration 20 mM and following compound 1 (compound of formula IIIa) which activates GK of final concentration 0.3 mM, and this was used for crystallisation. To protein solution 1-5 μl was added as crystallisation solution, an equal quantity of 28-30 % PEG 1500 and 0.1 M Hepes-NaOH (pH 6.6), and this solution formed by admixing was placed in a closed container containing 0.5-1 ml of crystallisation solution as that both solutions did not form contact, and the container was left to stand at 20°C. After standing for about 3 days - 1 months, crystals with maximum size of about 0.4 mm x 0.4 mm x 0.7 mm was obtained in the sample solution (Example 1).

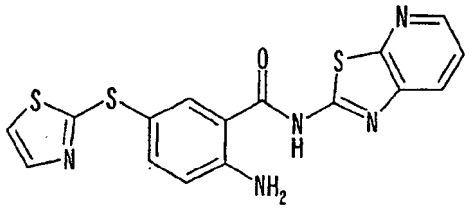
Furthermore, the crystals obtained by aforesaid method were immersed for about 3-7 days in 28-80 % PEG 1500, 0.1 M HePes-NaOH (pH 6.6) solution so that the following compound 2 (compound represented by formula IIIb) was contained in a concentration of 0.3 mM, and thereby a complex crystals of the following compound 2 and aforesaid mutant type GK were obtained.

Compound 1



(IIIa)

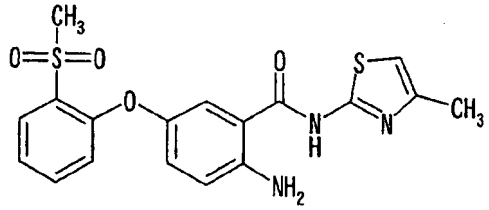
Compound 2



(IIIb)

Moreover, crystallisation was carried out in the same way as in Example 1 except that compound 3 (compound represented by formula IIIc) was used in stead of aforesaid compound 1, and as a result, crystals were respectively obtained in the same way as in Example 1 (Example 3).

Compound 3



(IIIc)

The obtained crystals were immersed into a crystallisation solution added with 10 % glycerol, thereafter it was rapidly frozen in liquid nitrogen. The X-ray diffraction data of the frozen crystal was collected in 100 K nitrogen gas stream by oscillation method at BL6B of synchrotron institution KEK-PF. From the obtained diffraction pattern, diffraction intensity was numerised using DENZO/SCALEPACK (made by HKL Co.), and crystal structure factor was determined. At this step, the crystal was found to be hexagonal system and had a space group of $P6_522$ or $P6_122$, and crystalline unit lattice was a = b = 79.9 angstrom, c = 322.2 angstrom, alpha = beta = 90°, gamma = 120°.

Using the obtained structure factor and three-dimensional structural coordinates of Human hexokinase type 1, the structure was analysed by molecular replacement method. Data with the resolution of 8 angstrom to 4 angstrom was used for the calculation, and it was performed by Amore program of CCP4 (Council for the Central laboratory of the Research Councils). The R factor of structure obtained by calculation was 53.7 %, and it was found that the space group of the crystal was $P6_522$ and a single molecule of mutant type GK was contained in asymmetrical unit. Electron density map was obtained

from this structure and structure factor, and the structure of mutant type glucokinase was determined using a program O (made by Dat-ONO Company).

Thereafter, refinement of the position of amino acid was carried out using CNX (Accelrys Inc.) and identification of amino acid residue was carried out using program O. This operation was repeated, and the structural coordinate of 448 amino acid residues from threonine 14 of the mutant type glucokinase to cysteine 461, 1 molecule of glucose molecule, 1 molecule of compound A, 1 sodium ion and 149 water molecules were identified, and the structural coordinates were determined. The R factor which is used as index of accuracy of finally determined structure was R = 23.2 % with respect to the data of resolution from 30 angstrom to 2.3 angstrom, and the R factor (Rfree) with respect to the data which was not used for the calculation in the refinement step of the structure was 27.4 %. There was no amino acid residue having the unacceptable structure by confirmation with Ramachandran plot.

The structure of the determined mutant type glucokinase was similar to the structure of the hexokinase which was isozyme, but the structure of the binding site of compound 1 (compound of formula IIIa) which activates glucokinase was different. This structural difference could not be expected with the ability of current computational chemistry and it became clear for the fist time by this structural analysis that this site was the binding site of activator and its detailed stereostructure. Figure 1a is ribbon diagram showing three-dimensional structure of the glucokinase elucidated here. As shown in Figure 1a, the newly found activator binding site was located between large domain and small domain, and it was about 20 angstrom away from the active center wherein glucokinase bonded with the substrate, glucose. The amino acid residue of glucokinase constituting the activator binding site was as follows. Tyrosine 61 - serine 69, glutamic acid 96 - glutamine 98, isoleucine 159, methionine 210 - tyrosine 215, histidine 218 - glutamic acid 221, methionine 235, arginine 250, leucine 451 - lysine 459.

Moreover, the binding scheme of compound 1 (compound of formula IIIa) with respect to this binding site is shown in Figure 2 and the structure of binding site of glucokinase is shown in Figure 3. The thiazole ring formed van der Waals contact with each amino acid side chain molecule of valine 62, valine 452, valine 455, and moreover the nitrogen atom on thiazole ring was hydrogen bonded with nitrogen atom of main chain of arginine 63. The nitrogen atom of amide on compound 1 was hydrogen bonded with oxygen atom of main chain of arginine 63. Benzene ring part of compound 1 was formed van der

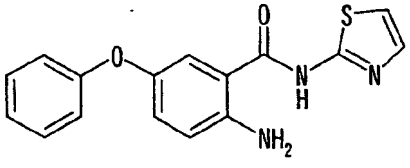
Waals contact with isolcucinc 211, and thc fluorinc atom substitutcd to bcnzcnc ring formcd van dcr Waals contacted with side chain of tyrosine 214. Aniline structure of compound 1 formed hydrogen bond with oxygen atom of side chain of tyrosine 215. Imidazole ring part bonded to the benzene ring via sulfur formed van der Waals contacted with amino acid side chain part of methionine 210, methionine 235, tyrosine 214. The serine 64-serine 69 part connecting the small domain and the large domain had a structure exposed to the solution, and compound 1 was bonded to the lower part of the arc-form structure formed by this part (Figure 3).

Example 4

Example of drug design

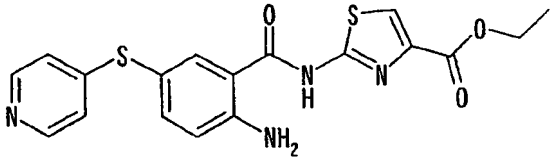
Using software UNITY (made by Tripos Company), pharmacophore of hydrogen bond acceptor and the hydrogen bond donor respectively generated from the main chain NII, CO of Arg 63, hydrophobic pharmacophore formed in the space corresponding to the phenyl group of aniline part of the ligand which formed the complex, and the protein surface formed on the basis of structure of the protein were used as search conditions, and Library compounds were screened, and the following compound 4 and compound 5 were obtained, and assay was carried out. As a result, activity of 780 % and 560 % was respectively observed. Wherein, the activity of 780 % denotes that the activity was enhanced upto 780 % by these compound when the activity of glucokinase of the control was 100 % (using glucose 2.5 M and ligand 10 μM).

Compound 4



Activity: 780 %

Compound 5



Activity: 560 %

Example 5

Crystal of mutant type GK (Δ1-15)

The crystal of simple substance of mutant type GK (Δ1-15) (glucokinase containing amino acid sequence represented by Sequence Number 8) was obtained using the following vapor diffusion technique.

In other words, mutant type GK purified to a high purity was concentrated, and finally a solution of mutant type GK of about 10 mg/ml (25 mM Tris-Cl pH 7.2, 50 mM NaCl, 5 mM TCEP) was formed. To protein solution 1-5 μl was added an equal quantity of crystallisation solution (1.5-1.6 M ammonium sulfate, 50 mM NaCl, 0.1 M Bicine NaOH (pH 8.7)), and this solution formed by admixing was placed in a closed container containing 0.5-1 ml of crystallisation solution as that both solutions did not form contact, and the container was left to stand at 20°C. After standing for about 3 days - 1 months, crystals with maximum size of about 0.07 mm x 0.07 mm x 0.5 mm was obtained in the sample solution.

The obtained crystals were immersed into the crystallisation solution added with 20 % glycerol, and continuing it was frozen rapidly in liquid nitrogen. The X-ray diffraction data of the frozen crystal were collected by oscillation method in 100K nitrogen gas stream at BL32B2 of synchrotron institution Spring-8. From the obtained diffraction image, the diffraction intensity was numerised using Mosflm, and crystal structure factor was determined. At this step, it became clear that the crystal was hexagonal system and had space group of $P6_522$ or $P6_122$, and crystal unit lattice was a = b = 103.2 angstrom, c = 281.0 angstrom, alpha = beta = 90°, gamma = 120°.

Next, molecular replacement method was carried out using the obtained structure factor and structure was analyzed. As model of stereostructure, three-dimensional structural coordinates of each domain of glucokinase determined from the mutant type GK (Δ1-11) / glucose / compound complex crystal was separately used. The calculation was performed by Amore program of CCP4 (Council for the Central laboratory of the Research Councils) using data of resolution of 8-4 angstrom. It was found that the space group of the crystal was $P6_522$, and a single molecule of mutant type GK (Δ1-15) was contained in the asymmetrical unit. Electron density map was obtained from this structure and structure factor, and the structure of mutant type GK (Δ1-15) simple substance was determined using program O (made by Dat-ONO Company).

Next, refinement of position of amino acid was carried out using CNX (made by Molecular Simulation Company) and identification of amino acid residue was carried out using program O. This operation was carried out repeatedly, and the structure coordinate of 424 amino acid residues from asparagine 180 to cysteine 461 and from methionine 15 to histidine 156 of mutant type glucokinase, and 2 molecules of sulfate ions, 1 sodium ion and 7 water molecules were identified, and the structural coordinates were determined. The R factor which is used as index of accuracy of finally determined structure was R = 23.8 % with respect to data of resolution of 50-3.4 angstrom, and the R factor (Rfree) with respect to the data which was not used for the calculation in the refinement step of structure was 30.6 %. There was no amino acid residue having the unacceptable structure by confirmation with Ramachandran plot.

The ribbon diagram showing the structure of glucokinase (Δ1-11) / glucose / compound 1 and the ribbon diagram showing the structure of glucokinase (Δ1-15) simple substance are respectively shown in Figure 1a and Figure 1b. Moreover, the figure on the right is a rotated figure of the figure on the left. In the structure of determined mutant type GK (Δ1-15) simple substance, the structures of main parts of the large domain and the small domain were similar to the respective structures of glucokinase determined by mutant type GK (Δ1-11) / glucose / compound complex crystal, but relative position of two domains was greatly different. In mutant type GK (Δ1-15) simple substance structure, the main part of the small domain was rotated by about 99 degrees from position of small domain in mutant type GK (Δ1-11) / glucose / compound complex structure. Moreover, alpha 13 helix located at C terminal region of glucokinase which constituted the small domain in the mutant type GK (Δ1-11) / glucose / compound complex structure no longer constituted the small domain in the mutant type GK (Δ1-15) simple substance structure, and it was located at between two domains. Moreover, because both the activator binding site and binding site of substrate, glucose were present between two domains in the mutant type GK (Δ1-11) / glucose / compound complex structure, the structures of their sites were greatly changed in the newly determined structure. The amino acid residues that play an important role in enzyme activity did not form active site in the mutant type GK (Δ1-15) simple substance structure, and the structure of mutant type GK (Δ1-15) simple substance analysed here was a structure of inactive state of glucokinase. Moreover, the activator binding site had completely disappeared in the structure of mutant type GK (Δ1-15) simple substance. The structural change of glucokinase (rotation of domains about 99°) observed by the mutant type GK (Δ1-11) / glucose / compound complex structure and the mutant type GK (Δ1-15)

simple substance structure was far greater compared with the previously known structural change of hexokinase (rotation of domains about 12°), and it could not be expected with the ability of current computational chemistry and it became clear from this structure analysis for the first time.

Moreover, in order to hinder the structural change to the inactive form mutant type GK (Δ1-15) simple substance structure, by designing a compound that binds to the compound binding site indicated by the mutant type GK (Δ1-11) / glucose / compound complex structure, it became clear that activator of glucokinase could be designed.

Possible Applications in Industry

As described above, in accordance with this invention, crystal of the glucokinase protein which was difficult to crystallise in the prior art was obtained. The three-dimensional structural coordinates obtained by analysing this crystal structure can be suitably used in order to design compounds that bind to glucokinase. Moreover, because the compounds designed in this way bind to glucokinase, they can be used as therapeutic agent of disease involving the glucokinase activity (for example diabetes mellitus therapeutic agent) as glucokinase activator or inhibitor.

Patent Claims

1. A glucokinase protein characterised in being used for crystallisation.

2. A protein in accordance with Claim 1 comprising amino acid sequence in accordance with Sequence Number 5.

3. A crystal of protein comprising amino acid sequence in accordance with Sequence Number 5 or amino acid sequence substantially the same amino acid sequence thereof.

4. A crystal in accordance with Claim 3, wherein the said protein is glucokinase protein.

5. A crystal in accordance with Claim 3 comprising crystals of protein containing amino acid sequence in accordance with Sequence Number 5.

6. A crystal in accordance with Claim 3, wherein the lattice constant satisfies the following equations (1)-(4)

a = b = 79.9 +/- 4 Å (1)

c = 322.2 +/- 15 Å (2)

alpha = beta = 90° (3)

gamma = 120° (4)

7. A crystal in accordance with Claim 6, wherein the space group is $P6_522$.

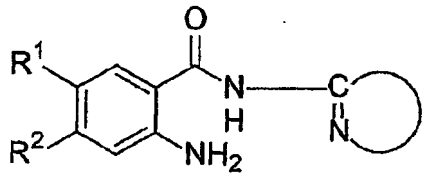
8. A crystal of protein specified by three-dimensional structure coordinates data in accordance with Table 1.

9. A crystal wherein in three-dimensional structure coordinates data changed in at least one data of three-dimensional structure coordinates data in accordance with Table 1, the mean square error between atoms of main chain of amino acid represented by three-dimensional structure coordinates data in accordance with Table 1 (C alpha atom) and C alpha atoms represented by the said changed three-dimensional structure coordinates data corresponding to aforesaid C alpha atoms is 0.6 Å or less.

10. A crystal in accordance with any of Claims 3-9, wherein the compound binding site is constructed by at least one of amino acid residues of tyrosine 61 - serine 69, glutamic acid 96 - glutamine 98, isoleucine 159, methionine 210 - tyrosine 215, histidine 218 - glutamic acid 221, methionine 235, arginine 250, leucine 451 - lysine 459 in amino acid sequence shown in sequence Number 5.

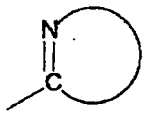
11. A crystal including a complex of the protein comprising amino acid sequence in accordance with Sequence Number 5 or amino acid sequence substantially the same amino acid sequence thereof and a compound which can bind to the said protein.

12. A crystal in accordance with Claim 11, wherein aforesaid compound is represented by formula (1).



(I)

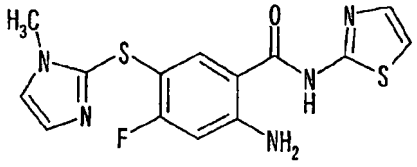
[wherein, R1 shows halogen atom, -S-(O)p-A, -S-(O)q-B or -O-B (wherein, p and q are the same or different and denote an integer of 0-2, A denotes C1-C6 alkyl group of optionally substituted straight chain, B denotes optionally substituted five-membered or six-membered ring aryl group or heteroaryl group, R2 denotes a hydrogen atom or halogen atom, and



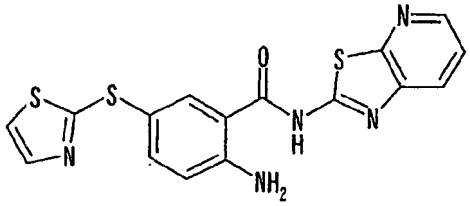
(II)

denotes an optionally substituted monocyclic or bicyclic heteroaryl group having a nitrogen atom adjacent to the carbon atom bonded to amide group].

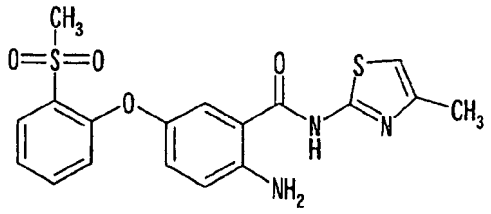
13. A crystal in accordance with Claim 12, wherein aforesaid compound is any of the compound represented by formula (IIIa)-(IIIc).



(IIIa)



(IIIb)



(IIIc)

14. A protein in accordance with Claim 1 comprising amino acid sequence in accordance with Sequence Number 8.

15. A crystal of protein comprising amino acid sequence in accordance with Sequence Number 8 or amino acid sequence substantially the same amino acid sequence thereof.

16. A crystal in accordance with Claim 15, wherein the said protein is glucokinase protein.

17. A crystal in accordance with Claim 15 comprising crystals of protein containing amino acid sequence in accordance with Sequence Number 8.

18. A crystal in accordance with Claim 15, wherein the lattice constant satisfies the following equations

a = b = 103.2 +/- 5 Å (5)

c = 281.0 +/- 7 Å (6)

alpha = beta = 90° (7)

gamma = 120° (8)

19. A crystal in accordance with Claim 18, wherein the space group is $P6_522$.

20. A crystal of protein specified by three-dimensional structure coordinates data in accordance with Table 2.

21. A crystal wherein in three-dimensional structure coordinates data changed at least one data of three-dimensional structure coordinates data in accordance with Table 2, the mean square error between atoms of main chain of amino acid represented by three-dimensional structure coordinates data in accordance

with Table 2 (C alpha atom) and C alpha atoms represented by the said changed three-dimensional structure coordinates data corresponding to aforesaid C alpha atoms is 0.6 Å or less.

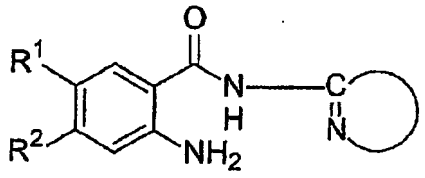
22. A process for the production of crystal containing a complex of protein and a compound that binds to the protein thereof, including

a protein production step wherein a protein containing the amino acid sequence having deletion of 1-50 amino acid residues from either or both of N terminal and C terminal of the protein containing amino acid sequence in accordance with Sequence Number 2 is produced, and

a protein reaction step wherein a compound that binds to the protein obtained in the said protein production step and the protein obtained in the said protein production step are reacted.

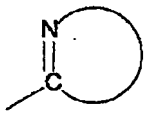
23. A process to produce crystal of the kind wherein a crystal of a protein is produced, characterised in that a protein including amino acid sequence in accordance with Sequence Number 5 or amino acid sequence substantially the same amino acid sequence thereof and having glucokinase activity and a compound which can bind to the said protein are used.

24. A process for the production of crystalline protein in accordance with Claim 23, wherein the compound which can bind to said protein is a compound represented by formula (1).



(I)

[wherein, R1 shows halogen atom, -S-(O)p-A, -S-(O)q-B or -O-B (wherein, p and q are the same or different and denote an integer of 0-2, A denotes C1-C6 alkyl group of optionally substituted straight chain, B denotes optionally substituted five-membered or six-membered ring aryl group or heteroaryl group, R2 denotes a hydrogen atom or halogen atom, and



(II)

denotes an optionally substituted monocyclic or bicyclic heteroaryl group containing nitrogen atom adjacent to the carbon atom bonded to amide group].

25. A process for the production of crystal in accordance with Claim 23 or 24 using co-crystallisation or soaking method

26. A drug design method of the kind wherein based on stereostructural information of a protein, the structure of compound that binds to said protein is designed, characterised in that the stereostructure information of said protein is the information obtained by analysing crystal in accordance with any of Claims 3-13 or 15-21.

27. A drug design method in accordance with Claim 26 characterised in that
a binding site deduction step wherein the compound binding site of said protein is deduced based on aforesaid stereostructure information, and
a selection step wherein a compound compatible to the compound binding site deduced in aforesaid binding site deduction step is selected from compound library,
are included.

28. A drug design method in accordance with Claim 26 characterised in that
a binding site deduction step wherein the compound binding site of said protein is deduced based on aforesaid stereostructure information, and
a compound structure assembly step wherein the structure of compound compatible to compound binding site deduced in aforesaid binding site deduction step is constructed,
are included.

29. A drug design method in accordance with Claim 26 characterised in that

a binding site deduction step wherein the compound binding site of said protein is deduced based on aforesaid stereostructure information, and

a design step wherein the structure of compound is designed by visual observation so that the compound binding site deduced in aforesaid binding site deduction step and a compound compatible to said compound binding site interact,

are included.

30. A drug design method in accordance with any of Claims 26-29, wherein aforesaid compound binding site is constituted by at least one of amino acid residue of tyrosine 61 - serine 69, glutamic acid 96 - glutamine 98, isoleucine 159, methionine 210 - tyrosine 215, histidine 218 - glutamic acid 221, methionine 235, arginine 250, leucine 451 - lysine 459 in amino acid sequence shown in sequence Number 5.

31. A drug design method in accordance with any of Claims 26-30 further including a step to measure physiological activity of the candidate compound predicted to be compatible to aforesaid compound binding site.

32. A drug design method in accordance with any of Claims 26-30 further including a binding determination step wherein the candidate compound predicted to be compatible to aforesaid compound binding site and a protein including amino acid sequence in accordance with and Sequence Number 5 or amino acid sequence which is substantially the same amino acid sequence thereof are contacted, and whether the candidate compound binds to the said protein or not is assessed.

33. A process for the production of compound array including the compound group selected by drug design method in accordance with any of Claims 26-30 is combined as compound array.

Rising Sun Communications Ltd. Terms and Conditions

Rising Sun Communications Ltd. shall not in any circumstances be liable or responsible for the accuracy or completeness of any translation unless such an undertaking has been given and authorised by Rising Sun Communications Ltd. in writing beforehand. More particularly, Rising Sun Communications Ltd. shall not in any circumstances be liable for any direct, indirect, consequential or financial loss or loss of profit resulting directly or indirectly from the use of any translation or consultation services by the customer.

Rising Sun Communications Ltd. retains the copyright to all of its' translation products unless expressly agreed in writing to the contrary. The original buyer is permitted to reproduce copies of a translation for their own corporate use at the site of purchase, however publication in written or electronic format for resale or other dissemination to a wider audience is strictly forbidden unless by prior written agreement.